Convolutional Approximations to the General Non-Line-of-Sight Imaging Operator

Byeongjoo Ahn[1] Akshat Dave[2] Ashok Veeraraghavan[2]
Ioannis Gkioulekas[1] Aswin C. Sankaranarayanan[1]

[1]Carnegie Mellon University [2]Rice University

Abstract

Non-line-of-sight (NLOS) imaging aims to reconstruct scenes outside the field of view of an imaging system. A common approach is to measure the so-called light transients, which facilitates reconstructions through ellipsoidal tomography that involves solving a linear least-squares. Unfortunately, the corresponding linear operator is very high-dimensional and lacks structures that facilitate fast solvers, and so, the ensuing optimization is a computationally daunting task. We introduce a computationally tractable framework for solving the ellipsoidal tomography problem. Our main observation is that the Gram of the ellipsoidal tomography operator is convolutional, either exactly under certain idealized imaging conditions, or approximately in practice. This, in turn, allows us to obtain the ellipsoidal tomography solution by using efficient deconvolution procedures to solve a linear least-squares problem involving the Gram operator. The computational tractability of our approach also facilitates the use of various regularizers during the deconvolution procedure. We demonstrate the advantages of our framework in a variety of simulated and real experiments.

1. Introduction

Recent advances in sensor technology have enabled high-speed imaging at picosecond timescales [32, 60]. This has facilitated *non-line-of-sight (NLOS) imaging* [32, 59], the ability of an imaging system to "look around corners" — a capability that finds application in assisted and autonomous driving, endoscopy, and imaging in confined spaces such as caves and debris.

The core ideas of NLOS imaging rely on measuring the time of flight (ToF) and radiance of multi-bounce photons that have interacted with the NLOS scene. The traditional measurement pipeline is as follows: a pulsed laser illuminates a LOS scene point and the ToF of photons arriving at a second LOS point is measured using a time-resolved sensor; these photons are assumed to have bounced off the NLOS scene via the LOS scene points. Binning the arriving photons based on their ToF provides a measurement of the light transport transient associated with the specific pair of LOS points. Repeating this process across multiple illumination and sensing points results in the measurement of the so-called 5D light transport transient [43, 49]. The 5D light transient provides a rich encoding of the geometry of the NLOS scene, thereby facilitating the recover of its shape.

Algorithmically, the reconstruction of the NLOS scene, modeled as a volumetric albedo map, from the 5D light transient requires the solution to an ellipsoidal tomography problem, which is usually computationally prohibitive to solve. The high computational cost can be attributed to two factors: first, the high-dimensionality of both the measurements and the NLOS scene voxelization; and second, the lack of structures in the measurement operator that facilitate fast implementations. Solving the linear inverse problem to recover a volumetric albedo from the 5D light transient measurements, with or without priors, is intractable.

Many existing methods avoid solving the inverse problem by relying on a pipeline adapted from computed tomography [28] called filtered backprojection. First, the backprojection operator is applied to the measured transients [2, 10, 18, 35, 59]; this operator projects each light transient measurement onto the voxels in the NLOS scene that could have contributed to it. Next, the Laplacian operator is applied to mitigate the enhancement of low-frequency components endemic to the backprojected result [10, 59]. Yet, this version of filtered backprojection as used in NLOS imaging is a heuristic reconstruction procedure, as it does not solve any specific formulation of the inverse problem.

Contributions. Our main technical result is to show that, under certain assumptions on the imaging geometry, the Gram[1] of the NLOS measurement operator is a *convolution operator*, as illustrated in Figure 1. This result advances NLOS imaging in three important ways. First, it allows us to efficiently obtain the ellipsoidal tomography reconstruction by solving an equivalent linear least-squares involving

[1]The Gram of the matrix $\mathbf{A}$ is $\mathbf{A}^\top\mathbf{A}$.

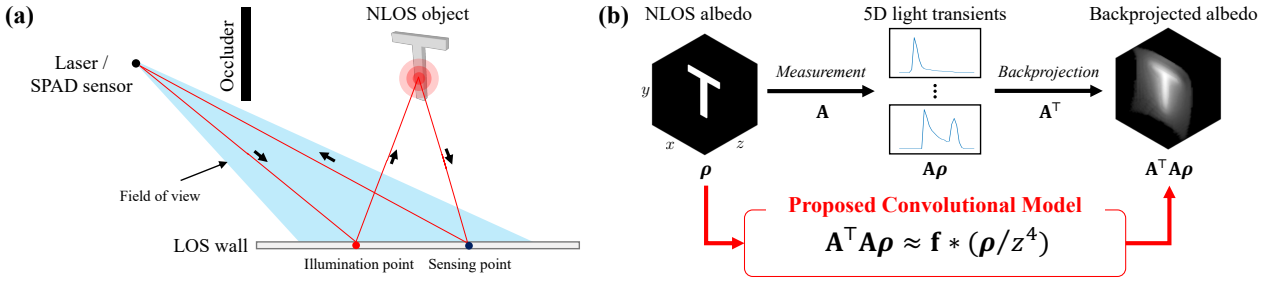


Figure 1. **Overview of our convolutional model for NLOS imaging.** (a) We capture transient measurements of light interacting with the NLOS scene through reflections on a relay LOS surface. (b) The measurement model $\mathbf{A}$ maps the volumetric albedo $\boldsymbol{\rho}$ of the NLOS scene to 5D light transients, and the backprojection operation $\mathbf{A}^\top$ maps 5D light transients to the backprojected albedo. We enable efficient and robust NLOS imaging by showing that the end-to-end mapping, modeled as the Gram operator $\mathbf{A}^\top\mathbf{A}$, is approximately convolutional.

the Gram operator: As the Gram operator is convolutional, this problem can be solved using computationally-efficient deconvolution algorithms. Second, it provides a theoretical justification for the filtered backprojection algorithm: We can show that filtered backprojection corresponds to using an approximate deconvolution filter to solve the problem involving the Gram operator. Third, it facilitates the use of a wide range of priors to regularize the NLOS reconstruction problem: The convolutional property of the Gram operator implies that the corresponding regularized least squares remain computationally tractable. We verify the benefits of this result using both simulated and real data.

2. Related Work

Non-line-of-sight imaging. Starting with the work of Kirmani et al. [31, 32], the past decade has seen a proliferation of works that explore different aspects of the non-line-of-sight (NLOS) imaging problem. We can broadly distinguish between two categories: Passive methods rely on environment illumination to localize or estimate rough motion and structure in the NLOS scene [4,5,7,8,13,52,55,56,63]. Active methods controllably inject light into the NLOS scene through a relay surface, which allows obtaining higher-fidelity information about it. For instance, even using only intensity cameras with coherent [6, 29, 30, 54] or incoherent illumination [11, 34] can enable accurate tracking and reconstruction of images of NLOS objects.

Our focus is on active methods that use light *transients*, that is, time-resolved radiometric measurements of the light reflected back by the NLOS scene [24]. These transients can be acquired using a variety of imaging technologies, including streak cameras [60], photonic mixer devices (PMD) [19,25,26,43], single-photon avalanche diodes (SPADs) [15], or interferometry [16]. For macroscopic scenes, SPADs have recently become the sensor of choice, providing time resolutions in tens of picoseconds [42, 53].

The technique closest to this work are those that recover discretized approximations of the NLOS scene by solving a problem of ellipsoidal tomography [10,20,21,27,35,44,45, 59]. Our paper serves to provide theoretical justification and computational acceleration for techniques commonly used within this framework. Alternative approaches for NLOS reconstruction from transient measurements use geometric algorithms based on Fermat's principle [57,62], wave-based models [37, 39, 50, 51], and inverse rendering [23, 47, 58]. These approaches provide different trade-offs in terms of reconstruction detail, reflectance invariance, and robustness to noise. Finally, some of these techniques have been extended to NLOS imaging with non-optical signals, including acoustic [36] and thermal [40].

Computed tomography (CT). NLOS imaging techniques based on ellipsoidal tomography are closely related to CT techniques [28]. In both settings, reconstruction can be reduced to solving a linear system, involving a measurement operator corresponding to different light trajectories inside an invisible volume: straight lines in the case of CT, and ellipsoidal shells in NLOS imaging. Because of this similarity, a lot of concepts from CT have found direct analogies in NLOS imaging, including the filtered backprojection algorithm [59] and analysis using the elliptic Radon transform [38]. Despite these analogies, our theoretical understanding of these concepts is often much more developed in the case of CT than in ellipsoidal tomography. The contributions of this paper help close this gap, for example by deriving the optimal filter to be used for filtered backprojection in the ellipsoidal case, analogous to the well-known Ram-Lak filter [28] in the linear case. This result is also of general interest for other problems where ellipsoidal tomography finds applications, for instance seismic imaging [41] and ultrasound imaging [3].

3. Problem Setup and Background

We begin by introducing the NLOS imaging problem, defining basic notation, and reviewing reconstruction procedures based on ellipsoidal tomography. We assume that we collect measurements through a *Lambertian* line-of-sight (LOS) surface $\mathcal{L}$. We use an impulse source (*e.g.*, pulsed laser) to illuminate a point $\mathbf{l}$ on $\mathcal{L}$. We then use a sensor

Table 1. **Comparison of NLOS imaging algorithms.** $\mathbf{A}$: Measurement operator. $\mathbf{i}$: 5D light transient. $\mathbf{L}$: Laplacian filter. $\mathbf{i}_c$: Confocal light transient ($\subset \mathbf{i}$). $\mathbf{k}_c$: 3D convolution kernel [44], $\mathbf{f}$: 3D convolution kernel (ours). m^2: Number of illumination (sensing) points. n_t: Number of time bins. n: Voxel resolution for each axis. Time complexity listed for iterative methods represent the cost per iteration.

Method	Formulation	Speed	Priors	Scanning pattern	Note	Complexity per iteration
Full linear reconstruction	$\operatorname{argmin}_{\boldsymbol{\rho}} \|\mathbf{i} - \mathbf{A}\boldsymbol{\rho}\|$	Slow	✓	No requirement	Intractable	$O(\min(m^4n^3, m^4n^2n_t))$
Backprojection	$\mathbf{A}^\top\mathbf{i}$	Fast	×	No requirement	Approximate	n.a.
Filtered backprojection	$\mathbf{L}\mathbf{A}^\top\mathbf{i}$	Fast	×	No requirement	Approximate	n.a.
Light-cone transform [44]	$\operatorname{argmin}_{\boldsymbol{\rho}} \|\mathbf{i}_c - \mathbf{k}_c * \boldsymbol{\rho}\|$	Very fast	✓	Confocal	Limited res.	$O(m^2n_t \log n_t)$
Ours	$\operatorname{argmin}_{\boldsymbol{\rho}} \|\mathbf{A}^\top\mathbf{i} - \mathbf{f} * (\boldsymbol{\rho}/z^4)\|$	Fast	✓	No requirement	Exact	$O(n^3 \log n)$

(*e.g.*, SPAD) to image a second point $\mathbf{s}$ on $\mathcal{L}$, and measure the *light transient*, *i.e.*, time-resolved intensity, $i(t; \mathbf{l}, \mathbf{s})$. The set of light transients at all pairs $\mathbf{l}$ and $\mathbf{s}$ on the LOS surface $\mathcal{L}$ is called the *5D light transient* [43, 49]:

$$I(\mathcal{L}) \equiv \{i(t; \mathbf{l}, \mathbf{s}) \mid \forall\, \mathbf{l}, \mathbf{s} \in \mathcal{L}\}. \tag{1}$$

Measurement model. To relate the 5D transient to the NLOS scene, we follow previous work [10, 18, 21, 32, 44, 59], and model the NLOS scene as a volumetric albedo function $\rho(\mathbf{x})$, where $\mathbf{x} = (x, y, z) \in \Omega_\mathbf{x}$ and $\Omega_\mathbf{x} \subset \mathbb{R}^3$ is the NLOS volume including the objects we are interested in recovering. Implicit in this model are the following assumptions: (1) each light path only interacts with a single NLOS scene point (three-bounce paths); (2) all NLOS scene points are visible from the LOS surface; and, (3) shading effects due to reflectance and normals are ignored, *i.e.*, light scatters isotropically at each NLOS scene point. For simplicity, we omit the portions of light paths contained in the LOS scene, *i.e.*, light travel to/from the illumination and sensing points to the imaging system. With these assumptions, we can define a *measurement operator* mapping the NLOS albedo $\rho(\mathbf{x})$ to light transient $i(t; \mathbf{l}, \mathbf{s})$ as

$$i(t; \mathbf{l}, \mathbf{s}) = \iiint_{\Omega_\mathbf{x}} \rho(\mathbf{x}) \frac{\delta(\|\mathbf{x} - \mathbf{l}\| + \|\mathbf{x} - \mathbf{s}\| - ct)}{\|\mathbf{x} - \mathbf{l}\|^2 \|\mathbf{x} - \mathbf{s}\|^2} \,\mathrm{d}\mathbf{x}, \tag{2}$$

where c is the speed of light. Because of the Dirac delta term $\delta(\cdot)$ in the integrand, the points in $\Omega_\mathbf{x}$ that contribute to the light transient i at a specific time instant t must satisfy

$$\|\mathbf{x} - \mathbf{l}\| + \|\mathbf{x} - \mathbf{s}\| - ct = 0. \tag{3}$$

We recognize this as the definition of a 3D ellipsoid with foci $\mathbf{l}$, $\mathbf{s}$, and major axis length ct [46]. Because each measurement accumulates information about the scene along an ellipsoid, inverting this measurement operator is referred to as a problem of *ellipsoidal tomography* [59]. Discretizing the measurement operator of Equation (2), we obtain

$$\mathbf{i} = \mathbf{A}\boldsymbol{\rho}, \tag{4}$$

where: $\mathbf{i} \in \mathbb{R}_+^{n_l n_s n_t}$ is the vectorized representation of light transient, with n_l, n_s, and n_t being the number of illumination points, sensing points, and time bins, respectively; $\boldsymbol{\rho} \in \mathbb{R}_+^{n_x n_y n_z}$ is the vectorized representation of the NLOS volumetric albedo, with n_x, n_y, and n_z being the number of voxels along each coordinate axis; and $\mathbf{A} \in \mathbb{R}_+^{n_l n_s n_t \times n_x n_y n_z}$ is a matrix corresponding to the discretization of integrand terms in Equation (2) other than the albedo ρ. We refer to $\mathbf{A}$ as the *measurement matrix*.

Reconstructing the NLOS scene can be done by solving the linear system of Equation (4), *e.g.*, in the least-squares sense, for the NLOS volumetric albedo $\boldsymbol{\rho}$. Unfortunately, this is generally not computationally tractable, because of the prohibitively large size of $\mathbf{A}$ as well as the lack of specific structures that facilitate fast implementations. (We defer a detailed comparison of the computational complexity of various reconstruction algorithms to Section 4.)

Backprojection. To circumvent this computational burden, an estimate of the NLOS volumetric albedo can be instead formed using the *adjoint* of the measurement operator of Equation (2), often referred to as the *backprojection operator* in analogy with computed tomography. Given the 5D light transients, the backprojection operator is written as

$$\rho_{\text{bp}}(\mathbf{x}) = \int\!\!\cdots\!\!\int_{\Omega_{\mathbf{l},\mathbf{s},t}} i(t; \mathbf{l}, \mathbf{s}) \frac{\delta(\|\mathbf{x} - \mathbf{l}\| + \|\mathbf{x} - \mathbf{s}\| - ct)}{\|\mathbf{x} - \mathbf{l}\|^2 \|\mathbf{x} - \mathbf{s}\|^2} \,\mathrm{d}t \,\mathrm{d}\mathbf{s} \,\mathrm{d}\mathbf{l}. \tag{5}$$

The backprojection operator projects each light transient $i(t; \mathbf{l}, \mathbf{s})$ onto the points $\mathbf{x}$ satisfying the ellipsoidal constraint of Equation (3). Then, the *backprojected volumetric albedo* $\rho_{\text{bp}}(\mathbf{x})$ at $\mathbf{x}$ is the weighted accumulation of these projections, with weights corresponding to light fall-off.

Upon discretization, Equation (5) becomes

$$\boldsymbol{\rho}_{\text{bp}} = \mathbf{A}^\top\mathbf{i}, \tag{6}$$

where $\boldsymbol{\rho}_{\text{bp}} \in \mathbb{R}_+^{n_x n_y n_z}$ is the vectorization of the backprojected volumetric albedo, and the *backprojection matrix* is simply the transpose of the measurement matrix $\mathbf{A}$.

There are many variants of the backprojection operator that have appeared in the NLOS imaging literature [10]. We will be making use of a version that omits the light fall-off terms in the integrand of Equation (5). In the discretized setting, this alternative operator can be expressed as

$$\boldsymbol{\rho}_{\text{bp,alt}} = (\mathbf{A} > 0)^\top\mathbf{i} = (\mathbf{A} > 0)^\top\mathbf{A}\boldsymbol{\rho}, \tag{7}$$

where $(\mathbf{A} > 0)$ is a binary matrix that equals one at the support of $\mathbf{A}$, and zero elsewhere. Our theoretical and algorithmic results will extend to this alternative definition as well. It is also worth noting that this alternative backprojection has important connections time-delay focusing [45] and it can be shown that the operator $(\mathbf{A} > 0)^\top \mathbf{A}$ is the imaging operator associated with that approach.

Filtered backprojection. Given that the backprojection of Equation (6) is simply the transpose to the measurement operator, the backprojected volumetric albedo $\boldsymbol{\rho}_{\text{bp}}$ is different from the solution of a linear least-squares problem based on Equation (4). Empirically, the backprojected volumetric albedo often resembles a low-frequency approximation of the true albedo $\boldsymbol{\rho}$. Therefore, it has been common practice to use a high-frequency filter as post-processing, to sharpen the backprojection results [10, 18, 59]. The most widely-used filter is the Laplacian,

$$\rho_{\text{fbp}}(\mathbf{x}) = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \rho_{\text{bp}}(\mathbf{x}). \tag{8}$$

Even though filtered backprojection works well in practice, it raises two important concerns. First, there is no formal justification for the use of the Laplacian filter. Second, there is no principled way to incorporate priors in the reconstruction pipeline. In many ways, our main contributions are to alleviate these concerns through mathematical analysis of the measurement and backprojection operators underlying NLOS imaging.

Confocal scanning. Before introducing our technical results, it is worth discussing a particular acquisition setting that has been termed *confocal scanning* by O'Toole *et al*. [44]. This corresponds to the case where only a three-dimensional subset of the 5D light transient is available, namely, the measurements corresponding to collocated illumination and sensing points, $\mathbf{l} = \mathbf{s}$. Then, O'Toole *et al*. [44] used the *light cone transform* (LCT) to show that the forward model of Equation (4) simplifies to:

$$\mathbf{i}_c = \mathbf{k}_{\text{LCT}} * \boldsymbol{\rho}, \tag{9}$$

where $\mathbf{i}_c$ is the vectorized representation of the confocal 3D light transient, and $\mathbf{k}_{\text{LCT}}$ is a 3D convolution kernel[2].

Reducing the measurement matrix $\mathbf{A}$ to a convolution enables recovering the volumetric albedo $\boldsymbol{\rho}$ by inverting (in the least-squares sense) Equation (9), , circumventing the need to use the heuristical filtered backprojection estimate. Further, the convolutional property makes it possible to regularize this inverse problem with priors such as non-negativity, sparsity and smoothness, without sacrificing efficiency. Inspired by this work, we aim to explore similar convolutional structure in the general non-confocal NLOS imaging scenario.

[2]The LCT also involves a non-linear reparameterization of all quantities along the z direction. We omit this for notational simplicity.

4. Convolutional Approximations

We are now ready to present our main theoretical and algorithmic results. For this, we begin with the following observation: The measurement model of Equation (4) suggests that we can recover the NLOS volumetric albedo $\boldsymbol{\rho}$ by solving the linear-least squares problem[3]:

$$\text{(P1)} \qquad \widehat{\boldsymbol{\rho}} = \underset{\boldsymbol{\rho}}{\operatorname{argmin}} \|\mathbf{i} - \mathbf{A}\boldsymbol{\rho}\|^2.$$

Provided $\mathbf{A}^\top \mathbf{A}$ is invertible, it is straightforward to show that $\widehat{\boldsymbol{\rho}}$ is also the solution to the problem:

$$\text{(P2)} \qquad \widehat{\boldsymbol{\rho}} = \underset{\boldsymbol{\rho}}{\operatorname{argmin}} \| \underbrace{\mathbf{A}^\top \mathbf{i}}_{\boldsymbol{\rho}_{\text{bp}}} - \mathbf{A}^\top \mathbf{A}\boldsymbol{\rho}\|^2,$$

where we recognize the first term in the square loss as the result of applying the backprojection matrix to the 5D transients (Equation (6)). In the rest of this section, we study the *Gram operator* $\mathbf{A}^\top \mathbf{A}$, and show that it can be well approximated as a convolutional operator. In turn, this facilitates the efficient solution of problem (P2).

4.1. The NLOS Gram operator

We first derive an expression for the continuous form of the Gram operator $\mathbf{A}^\top \mathbf{A}$. Note that the Gram operator maps an albedo volume $\boldsymbol{\rho}$ to the backprojected volume $\boldsymbol{\rho}_{\text{bp}}$,

$$\boldsymbol{\rho}_{\text{bp}} = \mathbf{A}^\top \mathbf{A}\boldsymbol{\rho}.$$

Substituting Equation (2) into Equation (5), we obtain:

$$\rho_{\text{bp}}(\mathbf{x}') = \iiint_{\Omega_{\mathbf{x}}} \rho(\mathbf{x}) \;\; k(\mathbf{x}, \mathbf{x}') \, \mathrm{d}\mathbf{x}, \tag{10}$$

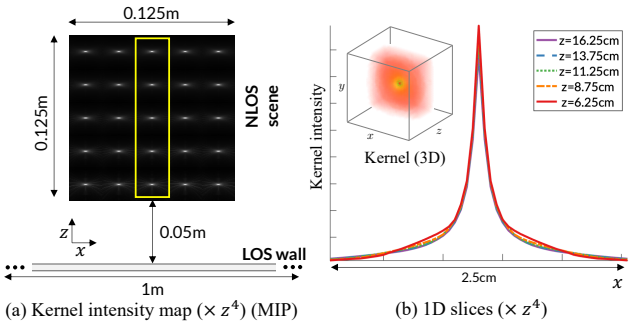
where $k(\mathbf{x}, \mathbf{x}')$ is a *spatially-varying kernel* that equals

$$\int\!\!\cdots\!\!\int_{\Omega_{\mathbf{l},\mathbf{s}}} \frac{\delta(\|\mathbf{x}'-\mathbf{l}\|+\|\mathbf{x}'-\mathbf{s}\|-\|\mathbf{x}-\mathbf{l}\|-\|\mathbf{x}-\mathbf{s}\|)}{\|\mathbf{x}-\mathbf{l}\|^2\|\mathbf{x}-\mathbf{s}\|^2\|\mathbf{x}'-\mathbf{l}\|^2\|\mathbf{x}'-\mathbf{s}\|^2} \, \mathrm{d}\mathbf{s}\, \mathrm{d}\mathbf{l}. \tag{11}$$

Gram operator as spatially-invariant blurring. We now state one of our main results, namely, that under certain assumptions on the LOS scene $\mathcal{L}$, the spatially-varying blur $k(\mathbf{x}, \mathbf{x}')$ can be expressed as $k(\mathbf{x}' - \mathbf{x})$. Then, the Gram operator is a convolution on the volumetric albedo.

Assumption 1 — Infinite planar LOS surface. Our first assumption is that the LOS scene $\mathcal{L}$ on which the illumination and sensing points lie is a *planar* surface of *infinite spatial extent.* In practice, this assumption holds approximately, so long as the NLOS scene, when projected orthographically on the LOS surface, is much smaller than the LOS surface. We discuss this in more detail later in this section.

[3]The use of square loss implies that our measurements have additive Gaussian noise. In practice, SPAD measurements are characterized by Poisson noise [22, 48]. As we show in experiments, our algorithms remain robust to this mismatch.



(a) Kernel intensity map ($\times z^4$) (MIP) (b) 1D slices ($\times z^4$)

Figure 2. **Spatial invariance of kernel k.** (a) Maximum intensity projection (MIP) of *normalized* ($\times z^4$) kernel intensity at 5×5 points. (b) 3D visualization and 1D slices of (a) at different depths. The kernels become spatially invariant after normalization.

Assumption 2 — Illumination and sensing points. Our second assumption is that the sensing and illumination points are continuously sampled on the LOS surface surface $\mathcal{L}$. In practice, a sampling density that allows for the integrals in Equation (11) to be well approximated would suffice.

Proposition 1 (Convolutional Gram operator) *Under Assumptions 1 and 2, the kernel $k(\mathbf{x}, \mathbf{x}')$ defined in Equation (11) can be written as*

$$k(\mathbf{x}, \mathbf{x}') = \frac{1}{z^4} f(\mathbf{x}' - \mathbf{x}), \tag{12}$$

where $\mathbf{x} = (x, y, z)$, and $\mathbf{x}'$ is in close vicinity to $\mathbf{x}$. The function $f(\cdot)$ is the kernel after normalization by z^4.

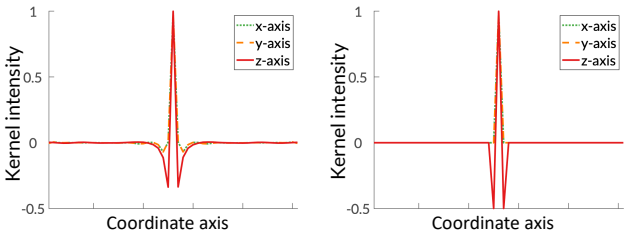
We provide the proof and analytical expression for $f(\cdot)$ in the supplement. Our proof uses the fact that, for an infinite LOS surface, every NLOS point has its corresponding set of illumination and sensing points where the ellipsoids have the same tangent planes. Thus, the kernel, which is the weighted superposition of ellipsoids, becomes spatially invariant after normalization by the $1/z^4$ term. This term is due to inverse-square light fall-off and we can absorb it inside $\rho(\mathbf{x})$ in Equation (10). Since we expect most of the energy of the kernel to be concentrated around $\mathbf{x}$, the kernel is largely spatially-invariant (see Figure 2). We can then rewrite Equation (10) as a 3D convolution:

$$\rho_{\text{bp}}(\mathbf{x}') = \iiint_{\Omega_x} \left(\frac{1}{z^4} \rho(\mathbf{x}) \right) f(\mathbf{x}' - \mathbf{x}) d\mathbf{x}. \tag{13}$$

Extension to alternative backprojection. Proposition 1 generalizes to the case when we use the alternative definition of backprojection in Equation (7) to define the Gram,

$$\boldsymbol{\rho}_{\text{bp, alt}} = (\mathbf{A} > 0)^\top \mathbf{A} \boldsymbol{\rho}. \tag{14}$$

In this case, the kernel is as in Equation (12) but *without* the $1/z^4$ term. We provide the proof in the supplement.



(a) Inverse kernel (ours) – 1D slice (b) Laplacian kernel – 1D slice

Figure 3. **Comparison of inverse kernel and Laplacian filter.** The plots show 1D slices of kernel values along each axis crossing through the center of the full 3D kernel.

Accommodating a finite LOS surface. When Assumption 1 is violated, that is, the LOS surface is finite, the blur kernel k of Equation (11) is spatially-varying, and therefore the Gram operator is not convolutional. However, when the NLOS scene is relatively small compared to the LOS surface, we expect that the spatially-varying kernel can still be well-approximated by a single convolutional operator. This is a common setting for all demonstrated NLOS imaging systems, which generally require LOS surfaces of much larger extent than the corresponding NLOS scenes.

We have verified empirically that, for typical NLOS scene and LOS surface sizes, the Gram matrix can be well approximated as a matrix of rank one. Additionally, we can use the top eigenvector of the matrix as a spatially-invariant kernel that better matches the non-ideal imaging geometry than the kernel k of Equation (12). We demonstrate this in the supplement, and we use this eigenvector-based kernel in all experiments in Section 5. In practice, this procedure can be used as a calibration step that has to be performed only once for each realization of the NLOS imaging geometry.

Inverse kernel. Figure 3 shows the *inverse* of kernel k, computed using Wiener deconvolution [61]. We observe that the inverse kernel closely resembles a Laplacian filter. This similarity lends theoretical support to the common choice of the Laplacian filter for post-processing in filtered backprojection [18, 59]. However, we must highlight two important differences. First, the discrete Laplacian filter is bereft of spatial scale. As a consequence, the result of traditional filtered backprojection is expected to change when we voxelize the NLOS scene at different resolutions. Our derived kernel and its inverse have no such adverse properties. Second, the inverse kernel in Figure 3 was derived under the two assumptions required for Proposition 1. When these assumptions do not hold, we can use eigendecomposition as discussed above to derive an approximate inverse kernel that can be significantly different from the Laplacian filter. Unlike filtered backprojection, our approach naturally accommodates for this by using the correct inverse filter.

4.2. Reconstruction with priors

Proposition 1 suggests a two-stage procedure for reconstructing the volumetric albedo $\boldsymbol{\rho}$: First, use backprojection to compute the backprojected volumetric albedo, $\boldsymbol{\rho}_{\text{bp}} = \mathbf{A}^\top \mathbf{i}$. Second, use deconvolution to solve the least-squares problem (P2) involving the Gram operator.

Proposition 1 additionally facilitates the solution of a more general version of problem (P2), namely,

$$\text{(P3)} \qquad \widehat{\boldsymbol{\rho}} = \operatorname*{argmin}_{\boldsymbol{\rho}} \| \underbrace{\mathbf{A}^\top \mathbf{i}}_{\boldsymbol{\rho}_{\text{bp}}} - \mathbf{A}^\top \mathbf{A} \boldsymbol{\rho} \|^2 + \Gamma(\boldsymbol{\rho}),$$

where the term $\Gamma(\boldsymbol{\rho})$ indicates a signal prior. Typical priors in NLOS imaging are positivity and sparsity both canonically as well as in spatial gradients (total variation) of the volumetric albedo $\boldsymbol{\rho}$ [20, 21]. State-of-the-art algorithms for solving regularized linear least-squares problems, such as alternating direction method of multipliers (ADMM) [9], are typically iterative, and thus require multiple forward and adjoint evaluations of the corresponding matrix—in the case of (P3), the Gram matrix $\mathbf{A}^\top \mathbf{A}$. These evaluations tend to dominate the cost of the algorithm. Therefore, the convolutional form of the Gram matrix offers a significant computational advantage, by turning forward and adjoint operations into simple convolutions. In the supplement, we derive the algorithmic details of using ADMM to solve (P3) with typical priors used for NLOS imaging.

Computational complexity. We can break down the cost of solving problems (P2) and (P3) into two parts. The first part is using backprojection to compute the backprojected volumetric albedo $\boldsymbol{\rho}_{\text{bp}} = \mathbf{A}^\top \mathbf{i}$. The time complexity of implementing $\mathbf{A}^\top$ (as well as $\mathbf{A}$) can be determined by the number of elements of the sparse matrix $\mathbf{A}^\top$, which is $n_l n_s n_x n_y n_z = m^4 n^3$ assuming $n_l = n_s = m^2$ and $n_x = n_y = n_z = n$. An alternative approach is to implement the operator by rendering ellipsoids corresponding to each time bin; the time complexity becomes $O(m^4 n^2 n_t)$. Thus, the time complexity of both the measurement operator and the backprojection is $O(\min(m^4 n^3, m^4 n^2 n_t))$.

The second part involves either one (for (P2)) or multiple (for (P3)) evaluations of the Gram operator $\mathbf{A}^\top \mathbf{A}$ and its adjoint. Using the convolutional approximation of the Gram operator, each such evaluation has complexity $O(n^3 \log n)$ using the 3D fast Fourier transform. Importantly, during the second part we do not need to perform any additional backprojection operations. In practice, the second part has negligible additional cost compared to the first part. Therefore, at a cost only marginally greater than filtered backprojection, our technique provides solutions to the full linear inverse problem, while also permiting the use of regularization.

By contrast, solving the linear inverse problem with or without priors using implementations of $\mathbf{A}$ and $\mathbf{A}^\top$ has per-iteration cost of $O(\min(m^4 n^3, m^4 n^2 n_t))$ [35], far greater than the combined cost of the two parts of our reconstruction procedure. Even though implementations of these operators can be accelerated using GPU implementations [23], our convolutional model reduces the inherent platform-independent computational complexity of the reconstruction procedure, enabling even more efficient implementations on an equivalent platform. Table 1 provides a summary of relative complexity of the various algorithms.

4.3. The confocal case

We prove in the supplement a version of Proposition 1 specifically for the case of confocal scanning, in which case the kernel of Equation (11) becomes

$$k_c(\mathbf{x}, \mathbf{x}') = \frac{1}{z^6} f_c(\mathbf{x}' - \mathbf{x}). \tag{15}$$

It is worth comparing the application of our approach to the confocal case with the LCT [44]. The two appear superficially similar, as they are both convolutional formulations for NLOS imaging. However, there is an important distinction: The LCT shows that (a non-linear reparameterization of) the measurement matrix $\mathbf{A}$ is convolutional, whereas our result shows the same for the Gram of the measurement matrix $\mathbf{A}^\top \mathbf{A}$. We discuss some implications of this difference.

Lateral resolution. The derivation of the LCT requires that the lateral (x and y) resolution of the NLOS volume be the same as the lateral resolution of collocated illumination/sensing points on the LOS surface; that is, $n = m$. Consequently, when the lateral scanning resolution is low, the lateral resolution of LCT results is necessarily poor. On the other hand, our method decouples NLOS volume resolution from the scanning resolution ($n \neq m$). Therefore, we can recover NLOS volumes at a lateral resolution that is often higher than the scanning resolution. The key here is to exploit the time resolution of light transients, which is often significantly higher than the lateral scanning resolution.

Computational complexity. Assuming equal volume lateral resolution and scanning resolution ($n = m$), LCT and our approach have the same computational complexity $O(m^2 n_t \log n_t)$ for evaluations of their corresponding operators (measurement and Gram, respectively), where $m^2 n_t$ is the number of elements in the confocal 3D transient.

However, we note an important disadvantage of our method: As discussed above, solving problem (P2) or (P3) requires performing a single backprojection operator, which has complexity (in the confocal case) $O(m^4 n_t)$. The LCT, on the other hand, operates directly on the light transients $\mathbf{i}_c$, and therefore has no such requirement.

Scanning pattern. The LCT can only be used if light transients are measured in a confocal scanning pattern. The confocal light transient provides rich information for NLOS imaging, and has advantages in terms of calibration

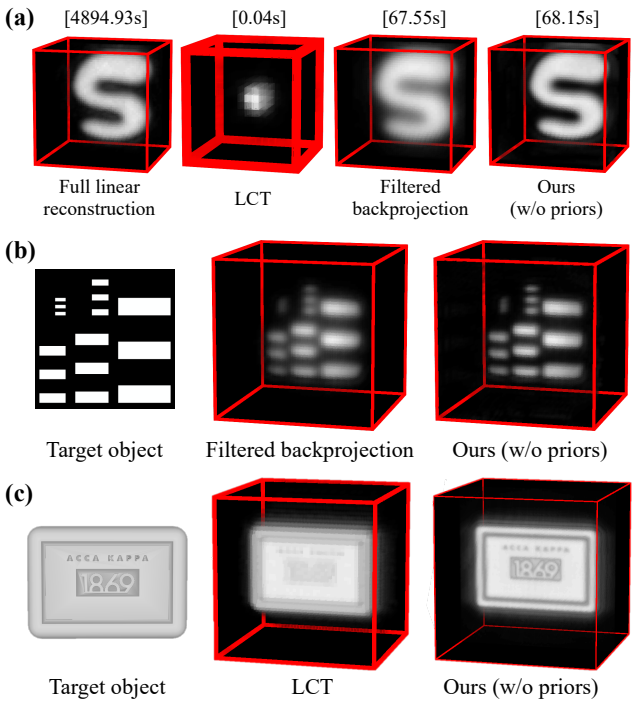


Figure 4. **Comparison under simulated 5D light transient.** Note that the thicker red lines in LCT indicates a lower resolution of the result caused by the inherent limit of the method.

and signal-to-noise ratio when measuring retroreflective objects [44]. However, confocal measurements can be corrupted by pile-up and after-pulsing effects [12, 17, 22, 48], due to the strong direct reflection from the LOS surface. By contrast, our approach is flexibly applicable to any scanning pattern, be it 3D non-confocal or 5D light transients.

5. Experiments

We compare our algorithm against full linear reconstruction, filtered backprojection, and LCT, on simulated and real transients. Although the convolutional model applies to both $\mathbf{A}^\top\mathbf{A}$ and $(\mathbf{A} > 0)^\top\mathbf{A}$, the $1/z^4$ term in the former causes numerical instability. Hence, we show results using the alternative backprojection. We provide implementation details and additional results in the supplement. Our implementation and data are available on our project page [1].

5.1. Simulated results

Rendering setup. For the simulation of light transient, we used light transients from the public NLOS imaging dataset [14, 33] as well as an implementation of a physically accurate renderer from [58]. The geometric parameters for each NLOS scene are provided in the supplement.

Comparisons with full linear reconstruction. Figure 4(a) compares our method and full linear reconstruction on scene consisting of an S-shape. Our method has faster runtime ($70\times$), while providing similar reconstruction quality.

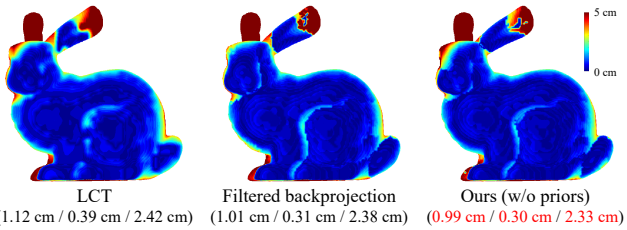


Figure 5. **Quantitative evaluation under simulated 5D light transient.** Error maps are shown with three error metrics (mean absolute error / median absolute error / root mean square error).

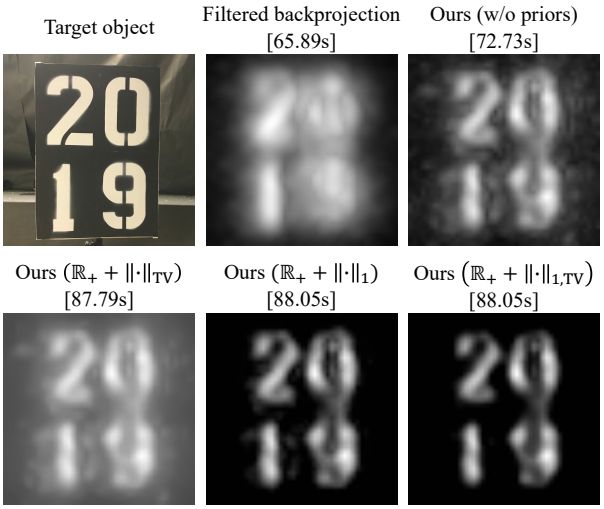


Figure 6. **Reconstruction with priors under real confocal light transient.** MIPs of recovered albedos when using priors. The priors enforce positivity, total variation, and sparsity of the volume.

Comparison with filtered backprojection. Figures 4(a) and 4(b) show comparisons between our method and filtered backprojection. Even without the use of priors, our method performs better, as it uses the exact inverse filter whereas filtered backprojection behaves as a heuristic inverse filter.

Comparison with LCT. Figure 4(c) shows the reconstruction under rendered confocal light transient. Here, while the temporal resolution of the light transient is $8\,\mathrm{ps}$, the coarse scanning of illumination points corresponds to $26\,\mathrm{ps}$ (*i.e.*, 64×64 samples on the LOS wall of size $0.5\,\mathrm{m} \times 0.5\,\mathrm{m}$). Thus, even though the temporal resolution of the light transient is enough to reconstruct the letters on the soap, it is not reconstructed well in LCT [44] due to the poor lateral resolution. On the other hand, we can recover significantly higher detail by running the same confocal measurements through our computational pipeline even without any prior.

Also, as shown in Figure 4(a), where LCT result is obtained from the confocal subset of the 5D light transient, LCT even cannot fully make use of the scanning resolution when the object is small compared to the LOS wall because the field of view is also tied to the range of LOS wall.

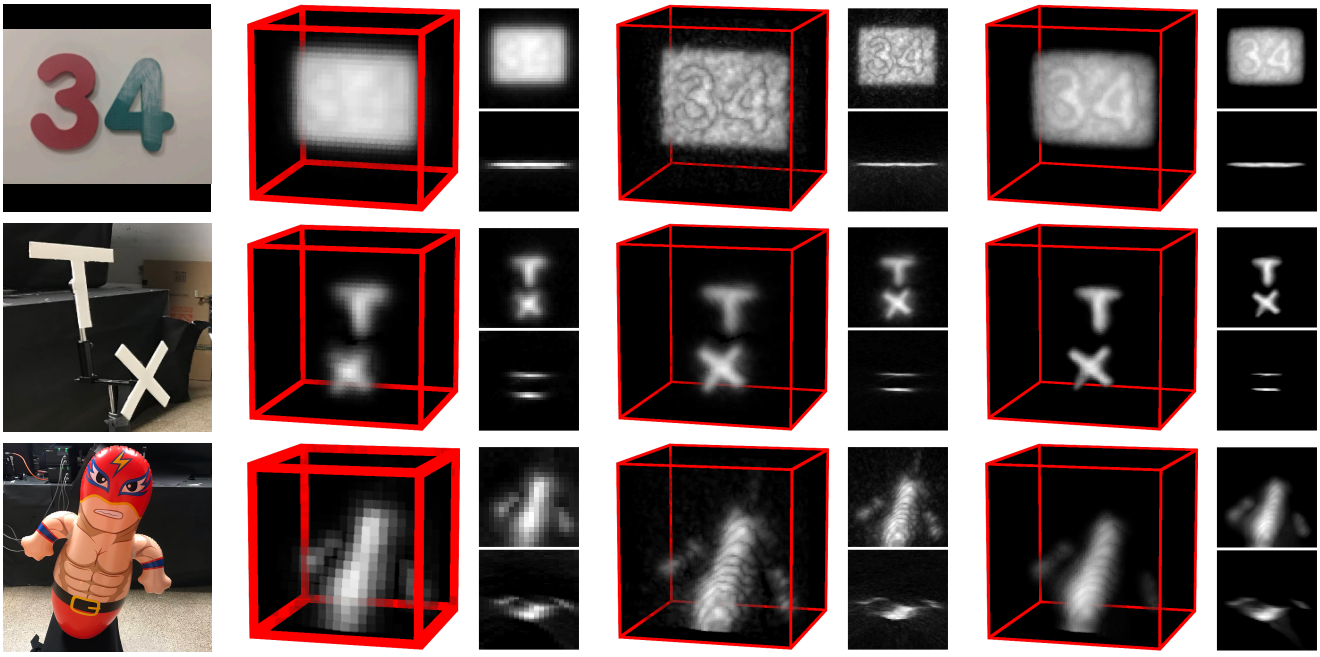


Figure 7. **Comparisons under real confocal light transients.** Our method produces reconstructions of higher detail and fewer artifacts than the alternatives. In our experiment, we cover the digits in the first row with white paper, to increase SNR.

Quantitative evaluation. Figure 5 shows point-wise differences between reconstructed MIP surface and the ground-truth surface along depth axis with three error metrics. We use 64×64 simulated confocal light transient for a scene of a Lambertian Stanford bunny. The temporal resolution of the measurements is $8\,\mathrm{ps}$, and the voxel resolution is $0.78\,\mathrm{cm}$ for LCT (inherent limit of the method) and $0.24\,\mathrm{cm}$ for FBP and our approach. We observe that our approach (without priors) performs the best on all error metrics.

5.2. Real scenes

Acquisition system. Our imaging system consists of a picosecond pulsed laser synced with a SPAD sensor. Details about the hardware prototype, calibration and acquisition procedures are in the supplement. Note that, although our method can be applied to the general 5D light transients as shown in Section 5.1, here we use confocal light transients because of their shorter acquisition time (1.5 hours for 51×51 points in 3D confocal, versus 11 hours for $12 \times 12 \times 12 \times 12$ points in 5D) and ease of calibration.

Reconstruction with priors. Figure 6 shows the effect of using different priors. We show results with a mixed prior that enforces positivity, total variation (*i.e.*, sparse gradients), as well as canonical sparsity enforced via ℓ_1 norm. We observe that, our method produces improvements over filtered backprojection even without priors, and that the use of priors further improves the reconstructions.

Comparisons. Figure 7 shows the comparisons with LCT and filtered backprojection under confocal light transients measured for various NLOS objects. We observe that our method provides higher reconstruction detail than LCT [44], by exploiting the high temporal resolution of transients, and fewer artifacts than filtered backprojection, by incorporating accurate inverse kernels and priors.

6. Conclusion

We presented convolutional approximations to the NLOS imaging operator by studying its Gram. This provides a pathway for applying priors of various kinds to regularize the inverse problem without a steep computational cost. Our method does not require a specific scanning pattern on LOS wall and can be adapted to exploit any specific subset of the 5D light transients. But above all, we believe that the insights we provide on the nature of the measurement operator for ellipsoidal tomography will have applications and impact on a range of problems, for instance seismic imaging, tissue imaging, and ultrasound imaging.

Acknowledgements. We thank Kiriakos Kutulakos and Srinivasa Narasimhan for helpful discussions. This work was supported by DARPA REVEAL (HR0011-16-C-0025, HR0011- 16-C-0028) and NSF Expeditions (CCF-1730574, CCF-1730147) grants. B. Ahn is supported by the Korea Foundation for Advanced Studies.

References

[1] Project page: Convolutional approximations to the general non-line-of-sight imaging operator. http://imaging.cs.cmu.edu/conv_nlos/, 2019. 7

[2] Victor Arellano, Diego Gutierrez, and Adrian Jarabo. Fast back-projection for non-line of sight reconstruction. *Optics Express*, 25(10):11574–11583, 2017. 1

[3] R Martin Arthur and Steven R Broadstone. Imaging via inversion of ellipsoidal projections of solutions to the linear acoustic wave equation. *IEEE transactions on medical imaging*, 8(1):89–95, 1989. 2

[4] Manel Baradad, Vickie Ye, Adam B Yedidia, Frédo Durand, William T Freeman, Gregory W Wornell, and Antonio Torralba. Inferring light fields from shadows. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 2

[5] Mufeed Batarseh, S Sukhov, Zhiqin Shen, Heath Gemar, Reza Rezvani, and Aristide Dogariu. Passive sensing around the corner using spatial coherence. *Nature communications*, 9(1):3629, 2018. 2

[6] Jacopo Bertolotti, Elbert G Van Putten, Christian Blum, Ad Lagendijk, Willem L Vos, and Allard P Mosk. Non-invasive imaging through opaque scattering layers. *Nature*, 491(7423):232, 2012. 2

[7] Jeremy Boger-Lombard and Ori Katz. Passive optical time-of-flight for non line-of-sight localization. *Nature Communications*, 10(1):3343, 2019. 2

[8] Katherine L Bouman, Vickie Ye, Adam B Yedidia, Frédo Durand, Gregory W Wornell, Antonio Torralba, and William T Freeman. Turning corners into cameras: Principles and methods. In *International Conference on Computer Vision (ICCV)*, 2017. 2

[9] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122, 2011. 6

[10] Mauro Buttafava, Jessica Zeman, Alberto Tosi, Kevin Eliceiri, and Andreas Velten. Non-line-of-sight imaging using a time-gated single photon avalanche diode. *Optics express*, 23(16):20997–21011, 2015. 1, 2, 3, 4

[11] Wenzheng Chen, Simon Daneau, Fahim Mannan, and Felix Heide. Steady-state non-line-of-sight imaging. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 2

[12] PB Coates. Pile-up corrections in the measurement of lifetimes. *Journal of Physics E: Scientific Instruments*, 5(2):148, 1972. 7

[13] Justin Dove and Jeffrey H Shapiro. Paraxial theory of phasor-field imaging. *Optics Express*, 27(13):18016–18037, 2019. 2

[14] Miguel Galindo, Julio Marco, Matthew O'Toole, Gordon Wetzstein, Diego Gutierrez, and Adrian Jarabo. Zaragoza nlos synthetic dataset, 2019. 7

[15] Genevieve Gariepy, Nikola Krstajić, Robert Henderson, Chunyong Li, Robert R Thomson, Gerald S Buller, Barmak Heshmat, Ramesh Raskar, Jonathan Leach, and Daniele Faccio. Single-photon sensitive light-in-fight imaging. *Nature communications*, 6:6021, 2015. 2

[16] Ioannis Gkioulekas, Anat Levin, Frédo Durand, and Todd Zickler. Micron-scale light transport decomposition using interferometry. *ACM Transactions on Graphics (TOG)*, 34(4):37, 2015. 2

[17] Anant Gupta, Atul Ingle, Andreas Velten, and Mohit Gupta. Photon-flooded single-photon 3d cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6770–6779, 2019. 7

[18] Otkrist Gupta, Thomas Willwacher, Andreas Velten, Ashok Veeraraghavan, and Ramesh Raskar. Reconstruction of hidden 3d shapes using diffuse reflections. *Optics express*, 20(17):19096–19108, 2012. 1, 3, 4, 5

[19] Felix Heide, Matthias B Hullin, James Gregson, and Wolfgang Heidrich. Low-budget transient imaging using photonic mixer devices. *ACM Transactions on Graphics (TOG)*, 32(4):45, 2013. 2

[20] Felix Heide, Matthew O'Toole, Kai Zang, David B Lindell, Steven Diamond, and Gordon Wetzstein. Non-line-of-sight imaging with partial occluders and surface normals. *ACM Transactions on Graphics (ToG)*, 38(3):22, 2019. 2, 6

[21] Felix Heide, Lei Xiao, Wolfgang Heidrich, and Matthias B Hullin. Diffuse mirrors: 3d reconstruction from diffuse indirect illumination using inexpensive time-of-flight sensors. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014. 2, 3, 6

[22] Quercus Hernandez, Diego Gutierrez, and Adrian Jarabo. A computational model of a single-photon avalanche diode sensor for transient imaging. *arXiv preprint arXiv:1703.02635*, 2017. 4, 7

[23] Julian Iseringhausen and Matthias B Hullin. Non-line-of-sight reconstruction using efficient transient rendering. *arXiv preprint arXiv:1809.08044*, 2018. 2, 6

[24] Adrian Jarabo, Belen Masia, Julio Marco, and Diego Gutierrez. Recent advances in transient imaging: A computer graphics and vision perspective. *Visual Informatics*, 1(1):65–79, 2017. 2

[25] Achuta Kadambi, Jamie Schiel, and Ramesh Raskar. Macroscopic interferometry: Rethinking depth estimation with frequency-domain time-of-flight. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 893–902, 2016. 2

[26] Achuta Kadambi, Refael Whyte, Ayush Bhandari, Lee Streeter, Christopher Barsi, Adrian Dorrington, and Ramesh Raskar. Coded time of flight cameras: sparse deconvolution to address multipath interference and recover time profiles. *ACM Transactions on Graphics (TOG)*, 32(6):167, 2013. 2

[27] Achuta Kadambi, Hang Zhao, Boxin Shi, and Ramesh Raskar. Occluded imaging with time-of-flight sensors. *ACM Transactions on Graphics (ToG)*, 35(2):15, 2016. 2

[28] Avinash Kak and Malcolm Slaney. *Principles of computerized tomographic imaging*. IEEE press New York, 1988. 1, 2

[29] Ori Katz, Pierre Heidmann, Mathias Fink, and Sylvain Gigan. Non-invasive single-shot imaging through scattering

layers and around corners via speckle correlations. *Nature photonics*, 8(10):784, 2014. 2

[30] Ori Katz, Eran Small, and Yaron Silberberg. Looking around corners and through thin turbid layers in real time with scattered incoherent light. *Nature photonics*, 6(8):549, 2012. 2

[31] Ahmed Kirmani, Tyler Hutchison, James Davis, and Ramesh Raskar. Looking around the corner using transient imaging. In *2009 IEEE 12th International Conference on Computer Vision*, pages 159–166. IEEE, 2009. 2

[32] Ahmed Kirmani, Tyler Hutchison, James Davis, and Ramesh Raskar. Looking around the corner using ultrafast transient imaging. *International journal of computer vision*, 95(1):13–28, 2011. 1, 2, 3

[33] Jonathan Klein, Martin Laurenzis, Dominik L Michels, and Matthias B Hullin. A quantitative platform for non-line-of-sight imaging problems. In *British Machine Vision Conference (BMVC)*, 2018. 7

[34] Jonathan Klein, Christoph Peters, Jaime Martín, Martin Laurenzis, and Matthias B Hullin. Tracking objects outside the line of sight using 2d intensity images. *Scientific reports*, 6:32491, 2016. 2

[35] Marco La Manna, Fiona Kine, Eric Breitbach, Jonathan Jackson, Talha Sultan, and Andreas Velten. Error backprojection algorithms for non-line-of-sight imaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2018. 1, 2, 6

[36] David B Lindell, Gordon Wetzstein, and Vladlen Koltun. Acoustic non-line-of-sight imaging. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 2

[37] David B Lindell, Gordon Wetzstein, and Matthew O'Toole. Wave-based non-line-of-sight imaging using fast fk migration. *ACM Transactions on Graphics (TOG)*, 38(4):116, 2019. 2

[38] Xiaochun Liu, Sebastian Bauer, and Andreas Velten. Analysis of feature visibility in non-line-of-sight measurements. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 2

[39] Xiaochun Liu, Ibón Guillén, Marco La Manna, Ji Hyun Nam, Syed Azer Reza, Toan Huu Le, Adrian Jarabo, Diego Gutierrez, and Andreas Velten. Non-line-of-sight imaging using phasor-field virtual wave optics. *Nature*, 2019. 2

[40] Tomohiro Maeda, Yiqin Wang, Ramesh Raskar, and Achuta Kadambi. Thermal non-line-of-sight imaging. In *IEEE International Conference on Computational Photography (ICCP)*, 2019. 2

[41] Sunghwan Moon and Joonghyeok Heo. Inversion of the elliptical radon transform arising in migration imaging using the regular radon transform. *Journal of Mathematical Analysis and Applications*, 436(1):138–148, 2016. 2

[42] Matthew O'Toole, Felix Heide, David B Lindell, Kai Zang, Steven Diamond, and Gordon Wetzstein. Reconstructing transient images from single-photon sensors. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 2

[43] Matthew O'Toole, Felix Heide, Lei Xiao, Matthias B Hullin, Wolfgang Heidrich, and Kiriakos N Kutulakos. Temporal frequency probing for 5d transient analysis of global light transport. *ACM Transactions on Graphics (TOG)*, 33(4):87, 2014. 1, 2, 3

[44] Matthew O'Toole, David B Lindell, and Gordon Wetzstein. Confocal non-line-of-sight imaging based on the light-cone transform. *Nature*, 555(7696):338, 2018. 2, 3, 4, 6, 7, 8

[45] Adithya Pediredla, Akshat Dave, and Ashok Veeraraghavan. SNLOS: Non-line-of-sight scanning through temporal focusing. In *IEEE International Conference on Computational Photography (ICCP)*, 2019. 2, 4

[46] Adithya Pediredla, Ashok Veeraraghavan, and Ioannis Gkioulekas. Elliptic path sampling for time-gated rendering. *ACM Transactions on Graphics (TOG)*, 3, 2019. 3

[47] Adithya Kumar Pediredla, Mauro Buttafava, Alberto Tosi, Oliver Cossairt, and Ashok Veeraraghavan. Reconstructing rooms using photon echoes: A plane based model and reconstruction algorithm for looking around the corner. In *IEEE International Conference on Computational Photography (ICCP)*, 2017. 2

[48] Adithya K Pediredla, Aswin C Sankaranarayanan, Mauro Buttafava, Alberto Tosi, and Ashok Veeraraghavan. Signal processing based pile-up compensation for gated single-photon avalanche diodes. *arXiv preprint arXiv:1806.07437*, 2018. 4, 7

[49] Raskar Ramesh and James Davis. 5d time-light transport matrix: What can we reason about scene properties? *MIT Technical Report*, 2008. 1, 3

[50] Syed Azer Reza, Marco La Manna, and Andreas Velten. Imaging with phasor fields for non-line-of sight applications. In *Computational Optical Sensing and Imaging*, pages CM2E–7. Optical Society of America, 2018. 2

[51] Syed Azer Reza, Marco La Manna, and Andreas Velten. A physical light transport model for non-line-of-sight imaging applications. *arXiv preprint arXiv:1802.01823*, 2018. 2

[52] Charles Saunders, John Murray-Bruce, and Vivek K Goyal. Computational periscopy with an ordinary digital camera. *Nature*, 565(7740):472, 2019. 2

[53] Dongeek Shin, Feihu Xu, Dheera Venkatraman, Rudi Lussana, Federica Villa, Franco Zappa, Vivek K Goyal, Franco NC Wong, and Jeffrey H Shapiro. Photon-efficient imaging with a single-photon camera. *Nature communications*, 7:12046, 2016. 2

[54] Brandon Smith, Matthew O'Toole, and Mohit Gupta. Tracking multiple objects outside the line of sight using speckle imaging. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6258–6266, 2018. 2

[55] Matthew Tancik, Guy Satat, and Ramesh Raskar. Flash photography for data-driven hidden scene recovery. *arXiv preprint arXiv:1810.11710*, 2018. 2

[56] Christos Thrampoulidis, Gal Shulkind, Feihu Xu, William T Freeman, Jeffrey Shapiro, Antonio Torralba, Franco Wong, and Gregory Wornell. Exploiting occlusion in non-line-of-sight active imaging. *IEEE Transactions on Computational Imaging (TCI)*, 2018. 2

[57] Chia-Yin Tsai, Kiriakos N Kutulakos, Srinivasa G Narasimhan, and Aswin C Sankaranarayanan. The geometry of first-returning photons for non-line-of-sight imaging. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 2

[58] Chia-Yin Tsai, Aswin C Sankaranarayanan, and Ioannis Gkioulekas. Beyond volumetric albedo–a surface optimization framework for non-line-of-sight imaging. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 2, 7

[59] Andreas Velten, Thomas Willwacher, Otkrist Gupta, Ashok Veeraraghavan, Moungi G Bawendi, and Ramesh Raskar. Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nature communications*, 3:745, 2012. 1, 2, 3, 4, 5

[60] Andreas Velten, Di Wu, Adrian Jarabo, Belen Masia, Christopher Barsi, Chinmaya Joshi, Everett Lawson, Moungi Bawendi, Diego Gutierrez, and Ramesh Raskar. Femto-photography: capturing and visualizing the propagation of light. *ACM Transactions on Graphics (TOG)*, 32(4):44, 2013. 1, 2

[61] Norbert Wiener. Extrapolation, interpolation, and smoothing of stationary time series: with engineering applications. *MIT press*, 1949. 5

[62] Shumian Xin, Sotiris Nousias, Kiriakos N Kutulakos, Aswin C Sankaranarayanan, Srinivasa G Narasimhan, and Ioannis Gkioulekas. A theory of fermat paths for non-line-of-sight shape reconstruction. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 2

[63] Feihu Xu, Gal Shulkind, Christos Thrampoulidis, Jeffrey H Shapiro, Antonio Torralba, Franco NC Wong, and Gregory W Wornell. Revealing hidden scenes by photon-efficient occlusion-based opportunistic active imaging. *Optics express*, 26(8):9945–9962, 2018. 2